(19)日本国特許庁（ＪＰ）　(12) **公開特許公報(A)**　(11)特許出願公開番号

特開平6－236195

(43)公開日　平成6年(1994)8月23日

(51)Int.Cl.[5]	識別記号	庁内整理番号	ＦＩ	技術表示箇所
Ｇ１０Ｌ　3/00	５１３　Ａ	9379－5H		

審査請求　未請求　請求項の数5　ＯＬ　（全８頁）

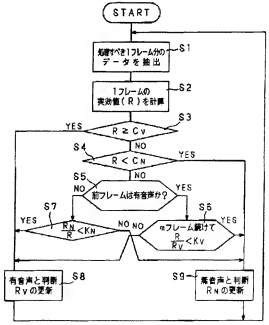
(21)出願番号　特願平5－24543

(22)出願日　平成5年(1993)2月12日

(71)出願人　000002185
ソニー株式会社
東京都品川区北品川６丁目７番35号
(72)発明者　南野　活樹
東京都品川区北品川６丁目７番35号　ソニー株式会社内
(74)代理人　弁理士　小池　晃　（外２名）

(54)【発明の名称】　音声区間検出方法

(57)【要約】
【構成】　入力音声信号から処理すべき１フレーム分のデータを抽出し（Ｓ１）、該１フレームの実効値（ＲＭＳ値）Ｒを計算し（Ｓ２）する。また、音声区間の最新の３０フレームのＲＭＳの平均値R_Tを計算しておき、各フレーム毎のＲＭＳ値ＲをR_Tで除算した値Ｒ／R_Tを求める。音声区間中は、この値Ｒ／R_Tが所定の閾値K_Tより小さくなることを検出し、これがm（例えば１５）フレーム連続するか否かを判別して（Ｓ６）、ＹＥＳのときには無音声区間の開始点とする（Ｓ９）。
【効果】　雑音のある環境下でも音声区間の検出を確実に行えるようにし、かつ演算量を少なくして電力消費を低減することができる。



1

【特許請求の範囲】

【請求項1】　入力された音声信号を一定サンプル数のフレームを単位として分割し、各フレーム毎に音声の有無を判別して音声の有る区間を検出する音声区間検出方法において、

音声区間とされた複数フレームの平均パワーにより各フレームのパワーを正規化する工程と、

この正規化された値を所定の閾値と比較して音声区間を検出する工程とを有することを特徴とする音声区間検出方法。

【請求項2】　無音声区間とされた複数フレームの平均パワーと各フレームのパワーとの比をとる工程と、

この比の値を他の所定の閾値と比較して上記音声区間の開始点を検出する工程とを有することを特徴とする請求項1記載の音声区間検出方法。

【請求項3】　上記閾値以下となるフレームが所定数以上連続したとき上記音声区間が終了したことを検出することを特徴とする請求項1又は2記載の音声区間検出方法。

【請求項4】　上記フレームのパワーが所定の無声音区間パワー閾値より小さいとき、当該フレームを無音声区間とする工程を有することを特徴とする請求項1、2又は3記載の音声区間検出方法。

【請求項5】　上記フレームのパワーが所定の有音声区間パワー閾値より大きいとき、当該フレームを有音声区間とする工程を有することを特徴とする請求項1、2、3又は4記載の音声区間検出方法。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、音声信号中の音声の有る区間を無音声区間と区別して検出する音声区間検出方法に関し、特に、音声符号化の前処理として音声区間を検出して無音声区間をゼロマスクする処理等に適用可能な音声区間検出方法に関する。

【0002】

【従来の技術】入力された音声信号を所定の音声符号化方式で符号化して伝送する場合（通信媒体を介して送信したり記録媒体に記録する場合等）において、符号化に先立って、入力信号中の音声が有る区間（有音声区間）と無い区間（無音声区間）とを区別しておき、無音声区間をゼロ信号でマスクする（ゼロマスクする）ような、いわゆるＶｏｘ制御あるいはＶｏｘ処理が知られている。

【0003】上記音声信号の符号化の具体的な例としては、ＭＢＥ（Multiband Excitation：マルチバンド励起）符号化、ＳＢＥ（Singleband Excitation:シングルバンド励起）符号化、ハーモニック（Harmonic）符号化、ＳＢＣ（Sub-band Coding:帯域分割符号化）、ＬＰＣ（Linear Predictive Coding：線形予測符号化）、あるいはＤＣＴ（離散コサイン変換）、ＭＤＣＴ（モデファイドＤＣＴ）、ＦＦＴ（高速フーリエ変換）等がある。

2

【0004】

【発明が解決しようとする課題】ところで、音声信号には背景雑音が含まれていることが多く、このため音声区間を切り出す際に、例えば無音声区間中に含まれたノイズと音声を区別することが困難である。

【0005】すなわち、例えば入力音声信号の実効値（いわゆるＲＭＳ、２乗平均根）を用いて音声の有無を検出する方法があるが、この場合、無音声区間であるにもかかわらず、環境雑音等のノイズが含まれていることによって有音声と判断してしまうという問題があり、音声とノイズとの区別が困難である。

【0006】また、信号に含まれている基本周期やピッチ等を検出したり、信号波形のゼロクロスの頻度を見たり、周波数成分の分布を見ること等を併用することで、音声区間検出の信頼性を高めることも考えられるが、処理が複雑で計算量が増大するという問題がある。これは、音声符号化装置や電話装置等の消費電力の増加につながり、電池駆動タイプの携帯用機器においては電池寿命の短期化という欠点に結び付くことになる。

【0007】本発明は、上記実情に鑑みてなされたものであり、音声の有無を雑音等に影響されることなく確実に判別でき、しかも処理が簡単で計算量も比較的小さくて済み、適用される機器の消費電力を節約することができる音声区間検出方法の提供を目的とする。

【0008】

【課題を解決するための手段】本発明に係る音声区間検出方法は、上記課題を解決するために、入力された音声信号を一定サンプル数のフレームを単位として分割し、各フレーム毎に音声の有無を判別して音声の有る区間を検出する音声区間検出方法において、音声区間とされた複数フレームの平均パワーにより各フレームのパワーを正規化する工程と、この正規化された値を所定の閾値と比較して音声区間を検出する工程とを有することを特徴としている。

【0009】この場合、１つあるいは２つの閾値を用いて音声区間から無音声区間への移行点（無音声区間の開始点）及び無音声区間から音声区間への移行点（音声区間の開始点）を求めるようにしてもよいが、この音声区間の開始点については、無音声区間とされた複数フレームの平均パワーと各フレームのパワーとの比をとり、この比の値を他の所定の閾値と比較して検出することが好ましい。

【0010】また、上記無音声区間の開始点については、上記正規化された値が上記閾値以下となるフレームが所定数以上連続したとき上記音声区間が終了したことを検出することが好ましい。

【0011】さらに、これらの音声区間の開始点検出及び無音声区間の開始点検出が誤検出となることを避ける

ために、上記フレームのパワーが所定の無声音区間パワー閾値より小さいとき当該フレームを無音声区間とし、上記フレームのパワーが所定の有音声区間パワー閾値より大きいとき当該フレームを有音声区間とすることが望ましい。

【0012】

【作用】ノイズが含まれる入力音声信号に対しても音声区間の検出が確実に行え、計算量も比較的少なくて済む。

【0013】

【実施例】以下、本発明に係る音声区間検出方法の好ましい実施例について、図面を参照しながら説明する。図1は、本発明の第1の実施例となる音声区間検出方法を説明するためのフローチャートである。

【0014】この図1において、入力されたディジタル音声信号に対して、ステップS1では処理すべきiフレーム分のデータが抽出され、次のステップS2でiフレームのパワーあるいは実効値、いわゆるRMS（2乗平均根）の値Rが計算される。次のステップS3では、上記実効値Rが所定の有音声区間パワー閾値C$_i$以上であるか否かを判別し、YESのときはステップS8に進み、NOのときはステップS4に進む。ステップS4では、上記実効値Rが所定の無音声区間パワー閾値C$_n$であるか否かを判別し、YESのときはステップS9に進み、NOのときはステップS5に進む。ステップS5では、時間的に前のフレームが有音声フレームか否かを判別し、YES（有音声）のときはステップS6に進み、NO（無音声）のときはステップS7に進む。

【0015】ステップS6では、音声区間と判別された最新の一定nフレーム（例えば30フレーム）のパワー（例えばRMS値）の平均値R$_i$を求めておき、音声フレーム毎のパワー（RMS値）Rを上記音声区間のパワー平均値R$_i$で正規化した値R／R$_i$を求め、この音声区間パワー平均値によるフレーム毎のパワー正規化値R／R$_i$について、一定フレーム数m（例えば15フレーム）以上続けて所定の閾値K$_i$（例えば0．1）よりも小さくなっているか否かを判別している。このステップS6でNOと判別されたとき、すなわち上記正規化値R／R$_i$が上記閾値K$_i$（例えば0．1）以上であるときにはステップS8に進み、YESと判別されたとき（R／R$_i$＜K$_i$のとき）にはステップS9に進む。

【0016】ステップS7では、無音声区間と判別された最新の一定nフレーム（例えば30フレーム）のパワー（例えばRMS値）の平均値R$_n$を求めておき、この無音声区間のパワー平均値R$_n$をフレーム毎のパワー（RMS値）Rで除算した（割り算した）値R$_n$／Rを求め、この除算値R$_n$／Rが所定の閾値K$_n$（例えば0．5）よりも小さいか否かを判別している。このステップS7でYESと判別されたとき（R$_n$／R＜K$_n$のとき）にはステップS8に進み、NOと判別されたときにはステップS9に進み、

【0017】ステップS8では、現在のフレームが有音声区間であると判断すると共に、上記音声区間のパワー平均値であるR$_i$を更新する。ステップS9では、現在のフレームが無音声区間であると判断すると共に、上記無音声区間のパワー平均値であるR$_n$を更新する。これらのステップS8あるいはS9の処理後に上記ステップS1に戻る。

【0018】以上のような音声区間検出方法の実施例によれば、音声信号にノイズが含まれていても、SN比がある程度大きい定常ノイズであれば、音声の有無を検出でき、しかも計算量は比較的小さいものとなっている。これにより、デジタル携帯電話等における音声信号の送信において、送信パワーを節約することができる。

【0019】次に、図2のAに示すような入力音声信号を、所定のサンプリング周波数f$_s$（例えば8kHz）でサンプリングし、図2のBに示すように所定サンプル数（例えば160サンプル）を単位として分割してそれぞれを1フレームとし、各フレームに音声が含まれるか含まれないかを検出するための操作の具体例について説明する。

【0020】ここで、前述したようないわゆるVox処理を行うフレームを図2のBに示すフレームとすると、この1フレーム160サンプルに時間的に連続する65サンプル先までの合計225サンプル（図2のC）の内の、最新の160サンプル（図2のD）を用いて上記有声音区間か無声音区間かの判定を行う。

【0021】この図2のDに示す判定フレーム（160サンプル）のサンプル値ついて、上記RMS（2乗平均根）の値を求め、これをRとする。図3は具体的な入力音声信号に対する上記RMS値の時間経過に伴う変化を示しており、横軸に時間経過をフレーム数で表し、縦軸に音声信号をパワーを上記RMS値で表している。この場合の入力音声信号は、音声レベルは標準的なレベルで、背景雑音なしのものを第1の音声信号試料として用いている。

【0022】一方、音声区間の最新のn（例えば30）フレームのRMS値の平均値を求めておき、これをR$_i$とする。同様に、無音声区間の最新のnフレームのRMS平均値も求めておき、これをR$_n$とする。

【0023】次に、各フレーム毎に、比R／R$_i$、R$_i$／Rを計算する。もし、背景雑音に比べて音声がある程度大きく（例えば、音声区間のRMS平均値が背景雑音のRMS平均値の10倍以上）、しかも背景雑音が定常であれば、

(1) 比R／R$_i$は音声区間では1．0近傍を変化し、無音声区間では0．0近傍を変化する。

(2) 比R$_i$／Rは無音声区間では1．0近傍を変化し、音声区間になるとその定常性が崩れる。

と考えられる。

5　　6

【0024】ここで図4及び図5は、上記図3に示したフレーム毎のRMS値が得られるような上記第1の音声信号試料が入力されるときのR比R／Rᵧ の値及び比Rᵧ／Rの値の時間変化を示している。

【0025】そこで音声区間中では上記比R／Rᵧ に着目し、この比R／Rᵧ がlよりある程度小さくなり、かつそれが一定区間続いたとき、例えば、R／Rᵧ ＜0．1、という条件がm（例えば15）フレーム以上続いたとき、を無音声区間の始まりとみなす。この閾値Kᵧ ＝0．1は、SN比20dB以上の背景雑音が存在しても、無音区間が検知できるようにするときの条件である。図4の具体例では、点aの時刻からR／Rᵧ が閾値Kᵧ ＝0．1を下回るようになり、これがmフレーム（15フレーム）続いた時点bが無音声区間の始まりとなる。

【0026】次に、無音声区間から音声区間への移行の検知は、上記比R／Rᵧ を他の閾値で併別して行うようにしてもよいが、本実施例では上記比Rᵪ ／Rの変化に着目して行っている。すなわち、無音声区間中では、上記比Rᵧ ／Rの定常性が崩れたとき、例えば、Rᵧ ／R＜0．5（＝Kᵪ ）、となったとき（瞬間）を音声区間の始まりとみなす。図5の具体例では、点aの時刻からRᵧ ／Rが閾値Kᵪ ＝0．5を下回り、この時点aが音声区間の始まりとなる。

【0027】さらに、これらの無音声区間の始まり検出や音声区間の始まり検出が、誤った検出となるのを避けるために、上記比R／Rᵧ 、Rᵧ ／Rの条件が満たされても、上記フレーム毎のRMS値がある閾値Cᵧ （例えば200程度）より大きなフレームは有音声区間とみなし、上記RMS値が他のある閾値Cᵪ （例えば、レベルの小さな音声のRMS平均値の1／20程度）より小さなフレームは無音声区間とみなす。

【0028】ここで、上記Cᵧ は上記有音声区間パワー閾値に相当し、従来において音声区間検出のために用いられていた閾値より大きい値とすることができる。すなわち本来の音声区間検出は上記R／Rᵧ を上記閾値Kᵧ で併別することにより行われ、上記閾値Cᵧ は誤検出防止のために設定されるものであって、確実に音声区間と判断できる程度の大きさとすればよいからである。また、上記Cᵪ は上記無声音区間パワー閾値に相当し、例えば音声があったとしても人の耳に聴こえない程度の値に設定すればよい。

【0029】ところで上記図3～図5は、入力音声信号として、音声レベルが標準で、背景雑音なしの第1の音声信号試料を用いた場合を示しているが、音声レベルが小さい場合や、背景雑音がある場合でも、音声区間の検出が確実に行える。

【0030】すなわち、図6は、音声レベルが小さく（－20dB）、背景雑音なしの第2の音声信号試料を入力信号としたときの各フレーム毎の上記RMS値を破線で示し、音声レベルは標準で、背景雑音あり（SN比26dB）の第3の音声信号試料を入力信号としたときの各フレーム毎のRMS値を実線で示している。この図6から明らかなように、各フレーム毎のRMS値だけでは上記第2の音声信号試料の音声区間と第3の音声信号試料の無音声区間とを区別する閾値が得られず、例えば第3の音声信号試料の無音声区間を音声区間と誤判定したり、第2の音声信号試料の音声区間を無音声区間と誤判定するような不具合が生じる。

【0031】これに対して、各信号の音声区間の最新のnフレームのRMS平均値で除算して正規化すると、図7、図8に示すようなグラフが得られる。すなわち、図7は上記第2の音声信号試料のフレーム毎のRMS値Rを、音声区間の最新の30フレームのRMS値の平均値Rᵧ で除算することで正規化した値R／Rᵧ を示しており、図8は上記第3の音声信号試料について同様な手順で正規化して得られた値R／Rᵧ を示している。

【0032】これらの図7、図8においては、所定の閾値Kᵧ （例えば0．1）により音声区間と無音声区間とを確実に区別することができる。ここで、上述した実施例と同様に、このR／Rᵧ の値を音声区間から無音声区間への移行点を検出するような用途に用いる場合には、音声区間中にR／Rᵧ が上記閾値Kᵧ ＝0．1を下回りかつこれが所定のm（例えば15）フレーム連続する時点を無音声区間の開始点とすればよい。図7の例では点aからmフレーム後、点bからmフレーム後、図8の例では点aからmフレーム後、点bからmフレーム後、等が上記無音声区間の開始点になり得る。ただし、上記所定数mを大きくとると各図の点aからmフレーム目以次の音声区間内になってR／Rᵧ が閾値Kᵧ ＝0．1を超えるため、無音声区間の開始点とはならなくなり、各図の点bからmフレーム目のみが無音声区間の開始点となる。

【0033】音声区間の開始点は、上記図7、図8のR／Rᵧ を他の所定の閾値で併別して検出してもよいが、上述したように、無音声区間の最新のnフレーム（例えば30フレーム）のRMS値の平均値Rᵧ を求めておき、各フレーム毎にR／Rを計算して、このRᵧ ／Rの値が所定の閾値Kᵧ （例えば0．5）を下回った時点を音声区間の開始点とすればよい。さらに、上述したように誤検出を防止するために、フレーム毎のRMS値を上記有音声区間パワー閾値Cᵧ や上記無声音区間パワー閾値Cᵪ で併別して、音声区間の始まりや無音声区間の始まりを検出するようにしてもよいことは勿論である。

【0034】このような実施例の音声区間の検出方法は、例えばディジタル携帯電話の音声圧縮動作の前処理に適用して好ましい。すなわち、一般に携帯電話装置は、屋外等の雑音のある環境下で使用されることも多く、音声区間の検出が重要とされるのみならず、本実施例の検出方法は計算量も比較的少なく、電力消費が少な

くて済み、送信パワーを節約することができ、電池寿命を長く保つことができる。

【0035】なお、本発明は上記実施例のみに限定されるものではなく、例えば、入力音声信号の1フレーム内のサンプル数や、RMS値の平均値（R_τやR$_\tau$）を得るためのフレーム数nや、無音声区間の始まりを検出するときのフレーム数m等は、上記具体的数値以外に任意に設定することができる。また、各閾値K_τ、K_δ、C_τ、C_δ等も上記具体例に限定されない。さらに、各フレームのパワーとしては、上記RMS（2乗平均根）値の代わりに、絶対値や、2乗値等を用いるようにしてもよい。

【0036】

【発明の効果】本発明に係る音声区間検出方法によれば、音声区間とされた複数フレームの平均パワーにより各フレームのパワーを正規化し、この正規化された値を所定の閾値と比較して音声区間を検出しているため、ノイズが含まれる入力音声信号に対しても音声区間の検出が確実に行え、計算量も比較的少なくて済む。従って、特にディジタル携帯電話装置等に適用した場合に、雑音のある環境下でも送信パワーを節約することができ、電池寿命を長く保つことができる。

【0037】また、音声区間の開始点については、無音声区間とされた複数フレームの平均パワーと各フレームのパワーとの比をとり、この比の値を他の所定の閾値と比較して検出することが好ましい。無音声区間の開始点については、上記正規化された値が上記閾値以下となるフレームが所定数以上連続したとき上記音声区間が終了したことを検出することが好ましい。さらに、これらの

音声区間の開始点検出及び無音声区間の開始点検出が誤検出となることを避けるために、上記フレームのパワーが所定の無声音区間パワー閾値より小さいとき当該フレームを無音声区間としたり、上記フレームのパワーが所定の有音声区間パワー閾値より大きいとき当該フレームを有音声区間とすることが好ましい。これらによって、音声区間検出の精度及び信頼性をより高めることができる。

【図面の簡単な説明】

【図1】本発明に係る音声区間検出方法の一実施例を説明するためのフローチャートである。

【図2】入力音声信号のフレーム区分を説明するための図である。

【図3】第1の音声信号試料についてのフレーム毎のRMS値を示すグラフである。

【図4】第1の音声信号試料についてのフレーム毎のRMS値Rを音声区間の最新の30フレームのRMSの平均値R$_\tau$で除算した値R／R$_\tau$を示すグラフである。

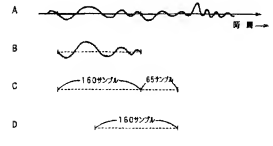
【図5】第1の音声信号試料についての無音声区間の最新の30フレームのRMSの平均値R$_\delta$をフレーム毎のRMS値Rで除算した値R$_\delta$／Rを示すグラフである。

【図6】第2の音声信号試料及び第3の音声信号試料についてのフレーム毎のRMS値を示すグラフである。

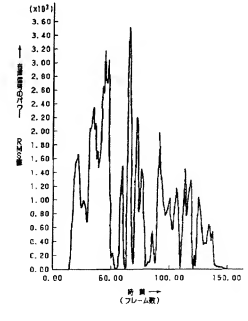
【図7】第2の音声信号試料についてのフレーム毎のRMS値Rを音声区間の最新の30フレームのRMSの平均値R$_\tau$で除算した値R／R$_\tau$を示すグラフである。

【図8】第3の音声信号試料についてのフレーム毎のRMS値Rを音声区間の最新の30フレームのRMSの平均値R$_\tau$で除算した値R／R$_\tau$を示すグラフである。

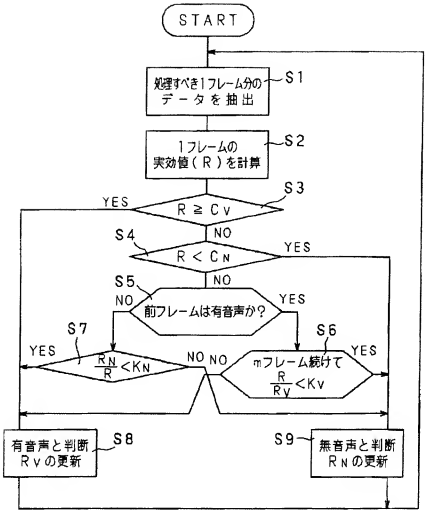
【図2】



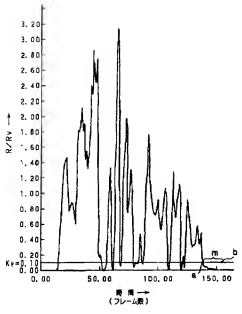
【図3】



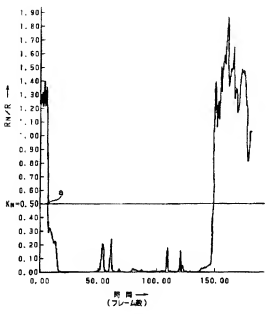
【図1】


START
処理すべき1フレーム分のデータを抽出
S1
1フレームの実効値(R)を計算
S2
S3
R ≧ CV
YES
NO
S4
R < CN
YES
NO
S5
NO
前フレームは有音声か?
YES
S7
YES
RN/R <KN
NO
NO
S6
mフレーム続けて
R/RV <KV
YES
有音声と判断
RVの更新
S8
S9
無音声と判断
RNの更新

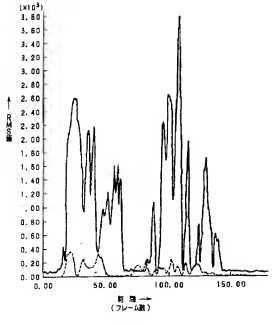
【図4】


R/Rv
Ky=0.10
a
b
m
0.00
50.00
100.00
150.00
時間
(フレーム数)

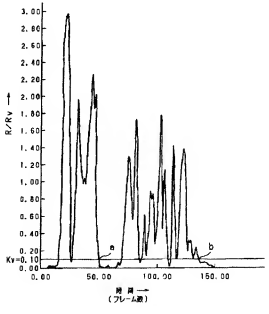
【図5】


R/Rv
Ka=0.50
a
0.00
50.00
100.00
150.00
時間
(フレーム数)

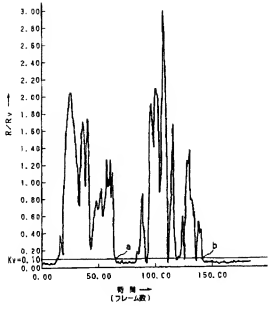
【図6】


(X10³)
0.00
50.00
100.00
150.00
時間
(フレーム数)

【図7】


R/Rv
Ky=0.10
a
b
0.00
50.00
100.00
150.00
時間
(フレーム数)

【図8】


3.00
2.80
2.60
2.40
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
R/Rv
Kv=0.10
a
b
0.00
50.00
100.00
150.00
時 間
(フレーム数)